DOCUMENT RESUME

ED 388 652 SP 036 300

AUTHOR Veenman, Simon; Raemaekers, Jan
TITLE Long-Term Effects of a Staff-Development Program on Effective Instruction and Classroom Management for Teachers in Multi-Grade Classes.
PUB DATE 31 Aug 95
NOTE 33p.; Paper presented at the European Conference for Research on Learning and Instruction (Nijmegen, Netherlands, August 26-31, 1995).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Classroom Techniques; Elementary Education; Foreign Countries; Inservice Teacher Education; Instructional Effectiveness; Longitudinal Studies; Maintenance; Mixed Age Grouping; *Multigraded Classes; *Staff Development; Teacher Effectiveness; Time on Task; Tutors
IDENTIFIERS *Long Term Effects; Netherlands

ABSTRACT

This paper describes the long-term effects of a staff-development program, "Dealing with Multi-Grade Classes," based on selected findings from teaching-effectiveness research in schools with multi-grade or mixed-age classes. The short-term effects were examined in two studies conducted in 1986-87 and 1989-90. In 1992 a follow-up study was conducted. A quasi-experimental, treatment-control group design was used to test the long-term effects of the program and the effects of coaching. Based on pre- and post-training classroom observations, the follow-up study revealed a significant treatment effect for the time-on-task levels of the pupils in the multi-grade classes and for the instructional and classroom management skills of the teachers. No significant differences were found between the coached and un-coached teachers and between the teachers who received the training program either 2 or 5 years ago. No significant differences were found between the post-test and retention test scores. This suggests that the training results were quite stable. No indication of further growth in the executive control of the selected instructional and classroom-management skills was found. No significant differences in achievement were found between the pupils in classes with trained teachers and the pupils in classes with untrained teachers. (Contains 33 references.) (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 388 652

Long-term Effects of a Staff-development Program on Effective Instruction and Classroom Management for Teachers in Multi-Grade Classes

Simon Veenman and Jan Raemaekers

Department of Educational Sciences,
University of Nijmegen,
The Netherlands

Paper presented at the European Conference for Research on Learning and Instruction

Nijmegen, August 26 to 31, 1995

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. Veenman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

SP036300

Abstract

This study describes the long-term effects of a staff-development program based on selected findings from teaching-effectiveness research in schools with multi-grade or mixed-age classes. The short-term effects of this program were examined in two studies directed at schools with multi-grade classes. The first improvement study was conducted in the school year 1986/87; the second improvement study was conducted in 1989/90. In the latter study, the effects of coaching in addition to participation in the staff-development program were also evaluated. In 1992, a retention or follow-up study was conducted. A quasi-experimental, treatment-control group design was used to test the long-term effects of the program 'Dealing with Multi-Grade Classes' and the effects of coaching. Based on pre-, post-training classroom observations, the follow-up study revealed a significant treatment effect for the time-on-task levels of the pupils in the multi-grade classes and for the instructional and classroom management skills of the teachers. No significant differences were found between the coached and uncoached teachers and between the teachers who followed the program either two or five years ago. No significant differences were found between the post-test and the retention test. This suggest that the training results were quite stable. No indication of further growth in the executive control of the selected instructional and classroom-management skills was found. No significant differences in achievement were found between the pupils in classes with trained teachers and the pupils in classes with untrained teachers.

In this study the long-term effects of a school improvement program directed at schools with multi-grade classes are reported. In two previous improvement studies, we assessed the short-term effectiveness of the staff-development program for teachers in multi-grade classes with respect to use of classroom time, instruction, and classroom management. The training program was inspired by the findings of our research into multi-grade classes (Veenman, Voeten, & Lem, 1987). The design of the training program was guided by research into effective staff development (Joyce & Showers, 1988). The results of these two studies have been published in *Educational Studies* (see Veenman, Lem, & Roelofs, 1989; Roelofs, Veenman, & Raemaekers, 1994). The purpose of the present study is to assess the effects of the staff-development program two to five years after training.

Background

The major impact of demographic contraction and staffing cuts since the middle 1970s on primary schools has been to increase the number of multi-grade classes. Multi-grade classes (also called mixed-age classes or combination classes) are classes in which pupils from two or more grades are taught by one teacher in one room at the same time. Pupils in multi-grade classes retain their respective grade-level assignments and follow their grade-specific curricula. These classes are generally formed for administrative and economic reasons. Schools confronted with either decreases or increases in pupil enrollment, for example, are forced to redistribute the pupils within the prescribed pupil-teacher ratios. Small schools in sparsely populated areas have always had multi-grade classes but the need for multi-grade teaching is now being faced by a much wider group of schools both in rural and urban areas.

In the Netherlands, 53% of the primary school teachers have a multi-grade class (Commissie Evaluatie Basisonderwijs, 1994). In a survey conducted in England and Wales, 40% of the schools surveyed reported an increase in multi-grade grouping as a result of falling enrollments (Walsh, Dunne, Stoten & Stewart, 1984). A further 15% reported that falling enrollments might lead to an increase in the extent of multi-grade teaching in the future. Almost one-half of the new teachers in England and Wales had their first appointment in multi-grade classes (Her Majesty's Inspectorate, 1982). One out of every seven classrooms in Canadian schools is a multi-grade classroom consisting of

two consecutive grades in one classroom. One out of every five pupils is enrolled in a multi-grade classroom in Canada. In the urban districts of this large country, moreover, a greater number of multi-grade classrooms are found than in the rural districts (Gayfer, 1991). The findings suggest that the multi-grade classroom occupies a significant position in our schools today.

Multi-grade teaching becomes a problem when it is forced upon the schools and the same groups cannot be maintained from year to year. Schools that are forced to set up multi-grade classes make greater demands on their teachers in terms of classroom organization and the creation of effective teaching learning conditions for the pupils. Based on the results of three observational studies and interviews with a number of teachers in multi-grade classes in the Netherlands, five problem areas have been identified: (1) the efficient use of instructional time, (2) the design of effective instruction, (3) classroom management, (4) the organization of independent study, and (5) agreement upon the goals of multi-grade teaching (Veenman, Lem, Voeten, Winkelmolen, & Lassche, 1986).

To assist teachers in multi-grade classes, a staff-development program was designed. This program was centred around the problem areas identified above and incorporated selected findings from previous research on teacher and school effectiveness. In designing this program, it was recognized that staff development or inservice activities often do not produce lasting effects (Van Tulder, 1992) and special attention was therefore devoted to the transfer of training.

Transfer of Training

A considerable amount of time, energy, and money is invested in staff development or inservice training today. Reviews of the literature on training, however, have indicated that little empirical attention has been devoted to the issue of training transfer (Baldwin & Ford, 1988; Broad & Newstrom, 1992). The instructional experiences provided by training are designed to develop new skills and new knowledge for application on the job. Transfer of training is defined as the degree to which the skills and knowledge acquired during training are effectively applied in the workplace. For transfer to occur, the trained behaviour must be generalized to the job context and maintained over a period of time.

Full transfer also means that the level of the skill increases with on-the-job practice to beyond the level demonstrated at the end of the training program.

Considering the low levels of transfer found for all types of training, Broad and Newstrom (1992) have assumed that perhaps 50% of training content may still be applied one year after training. In order to promote conditions for transfer, Broad and Newstrom propose the formation of transfer partnerships that include the trainees (i.e., learners), the trainers (i.e., designers and deliverers of the learning experiences), and the managers (i.e., leaders in the organization with the authority and responsibility for the application of the learning on the job). Each partner has an important contribution to make to the transfer process, and full transfer requires that all of the partners cooperate to maximize the application of the new skills and knowledge on the job. Each partner can utilize a number of strategies before, during, and after training to enhance the transfer. Broad and Newstrom (1992) developed an overview of the transfer strategies in the form of a matrix combining the time dimension (before-during-after training) with the role dimension (manager-trainer-trainee). Some of the strategies for managing transfer before training are, for example: collect baseline performance data, involve supervisors and trainees in needs analysis procedures (performed by the managers), systematic design instruction, involve managers and trainees (performed by the trainers), actively explore training options, participate in advance activities (performed by the trainees). A number of the strategies identified by Broad and Newstrom (1992) to facilitate the transfer of training have been incorporated into the design and execution of the staff-development program 'Dealing with Multi-Grade Classes: A Program for School Improvement'.

The staff development program

The following five topics have been considered in the program (Veenman, Lem, & Nijssen, 1988):

1. Instructional time. This topic is based on the notion that time is an essential element in learning and a potentially useful instructional variable. The way in which teachers and pupils spend their time provides valuable insights into the effectiveness of the teaching-learning process in multigrade classes. Results of the syntheses of several

thousand individual studies of academic learning conducted during the past half century in different countries show that instructional time has an overall correlation of about 0.4 to learning outcomes (Walberg, 1986; Fraser, Walberg, Welch, & Hattie, 1987). Teachers were informed about the importance of concepts such as pupil-engaged learning time, time needed for and spent in learning, time allocation, pupils' success levels, task appropriateness. Teachers were encouraged to use strategies that help pupils' stay-on-task. In addition, several observational methods were presented to observe pupils' time-on-task levels. Instructional time is an important topic for teachers in mixed-age classes because the complexity of the classroom organization may lead to lower levels of time-on-task.

2. *Effective instruction.* The research on effective teaching has yielded a pattern of instruction that is particularly useful for teaching a body of content or well-defined skills. In general, researchers have found that when effective teachers teach concepts and skills explicitly, they begin a lesson with a short statement of goals and a short review of previous, prerequisite learning. They present new material in small steps, provide active practice for all pupils, guide pupils during initial practice, provide feedback and correctives, and supervise pupils during seatwork or independent practice. Effective teachers also review in weekly and monthly intervals (Rosenshine & Stevens, 1986).

Teachers were informed of the findings of this research and of the key instructional behaviours as defined by Good, Grouws and Ebmeier (1983). They were encouraged to design lessons using these very specific components. Pupils in multigrade classes work more in an individual seatwork setting. In this setting, significantly less time is spent on the task as compared to the whole class or direct instruction setting. Important steps in the lesson plans for teachers in multigrade classes are guided and independent practice. After presentation of new material the teacher has to supervise the pupils' initial practice to make sure that they can practice independently with minimal difficulty when the teacher is instructing another group of pupils. At that moment the teacher is too busy to supervise the first group.

3. *Classroom management and organization.* Classroom management includes all the things teachers must do to foster pupil involvement and cooperation in classroom activities and to establish a productive working environment. Teachers were informed of ways to

manage their classes, largely in the light of research conducted by Kounin (1970) and Evertson, Emmer, Clements, Sanford and Worsham (1984). According to Kounin successful managers are aware of what is happening in classrooms (with-itness), are able to handle two or more simultaneous events (overlapping), to sustain a group focus (group alerting and accountability) and to keep the action moving along smoothly (smoothness and momentum). Based on the work of Evertson et al. (1984) teachers were informed of ways of organizing a good room arrangement, planning and using classroom rules and procedures, managing pupils' work and maintaining good pupils' behaviour. In multigrade classes teachers are tested more on their classroom management skills than teachers in single-age classes (Veenman, Voeten, & Lem, 1987). Teachers in multigrade classes with high levels of on-task behaviour were effective classroom managers. Their classes were well organized and well managed.

4. *Independent learning*. Pupils in multigrade classes spend most of their time in an independent seatwork setting. While one group of pupils is working individually, the teacher is teaching another group. Therefore, pupils in multigrade classes need to be adequately prepared during instruction. Teachers are informed of some instructional procedures that can help increase pupil engagement during seatwork, including the following: a) the teacher spends more time in demonstration (explaining, discussion) and guided practice, b) the teacher makes sure pupils are ready to work alone, by achieving a correct response rate of 80% or higher during guided practice, c) the seatwork activity follows directly after guided practice, d) the seatwork exercises are directly relevant to the demonstration and guided practice activities, e) the teacher guides the pupils through the first few seatwork problems (Rosenshine & Stevens, 1986). Attention is also given to the organization of multitasks, i.e. tasks in which pupils plan, select and organize materials and activities. In multi-task settings teachers are unable to control directly what each pupil is doing. In the program teachers were informed of ways to structure the working environment, largely in the light of Kierstead's work (1986). One aspect of the multi-task setting is the use of the pupils' work cycle; a set of routines, procedures, rules and consequences that spells out for pupils exactly what is expected of them: how they are to proceed and to account for the responsible use of their time.

5. School climate and school leadership. Teachers and their principals were given some results of the research on school effectiveness. In general terms the importance of cooperation, team spirit, shared values and norms and instructional school leadership was stressed. In our research we found that some teachers in multigrade classes felt very isolated from their colleagues working in single-age classes. Some outcomes of school effectiveness research highlighted factors such as school site management, active leadership, high expectation for pupils, change-supportive norms, school-wide staff development, clear goals, collaborative planning and collegial relationships (Levine & Lezotte, 1990; Good & Brophy, 1986). The content of this part of the program was not directed at changing teaching behaviours, but on stressing the importance of shared problem solving, peer support and a planned, purposeful program for dealing with mixed-age classes on a school-wide basis.

The contents of the program are integrated into a model for school and classroom effectiveness. This model comprises the following components: leadership, school climate, teacher behaviours, pupil behaviours and pupil achievement. Each chapter in the program contained a rationale, a definition of terms, and specific guidelines for the implementation of the instructional behaviours in multi-grade classes. To facilitate the understanding and use of the information in the program, numerous case studies were provided, along with several checklists. Videotapes were also developed to demonstrate some of the behaviours involved in effective teaching and classroom management. In the second school improvement study, moreover, coaching was added to facilitate the transfer of the training.

Coaching

Research on training effects has shown a frequent failure to transfer the new knowledge and skills or, when initial transfer has been accomplished, rapid attrition of the newly acquired behaviours. Few studies have actually measured the transfer effects of training, but recent analyses show transfer to only occur when in-class coaching has been added to the initial training experience that include theory, demonstration, practice, and feedback (Bennett, 1987; Joyce & Showers, 1988).

Coaching is defined by Joyce and Showers (1980) as: "Hands-on, in-classroom

assistance with the transfer and application of skills to the classroom." The process of coaching includes five major functions: (1) the provision of companionship, (2) the provision of technical feedback, (3) the analysis of application, (4) the adaptation to the pupils, and (5) personal facilitation. The first function is to provide interpersonal exchange with regard to a difficult process (i.e., the adoption of a new teaching strategy). This can result in mutual reflection, the checking of perceptions, the sharing of frustrations and successes, and thinking through mutual problems. The second function, the provision of technical feedback, helps ensure growth through practice in the classroom. Technical feedback includes pointing out omissions in the instructional strategy, examination of how the instructional materials are arranged, and checks for integration of the teaching strategy. The third function, the analysis of application, involves activities such as the selection of the appropriate occasions for the use of a newly acquired teaching strategy and examination of the existing curriculum for adequate use of the strategy. The fourth function, the adaptation to the pupils, involves learning how to teach the new strategy to the children. The fifth function, personal facilitation, refers to helping the teachers feel good about themselves during the early trials.

Implementation of the Staff-Development Program

In the first study directed at the evaluation of the short-term effectiveness of the staff-development program, the program was conducted by members of the Department of Educational Sciences at the University of Nijmegen for teachers of grades one through six. In the school year 1986/87, the staff-development program was followed by 41 teachers from 8 schools. From this group, 17 teachers were then selected for participation in the observational study (treatment group). Nine teachers from 6 schools in the same area were selected for observation but did not receive training (control group).

Based on the experiences of the teachers with the first version of the program, the contents were slightly revised for the second version. The kindergarten teachers felt that too little attention was paid to kindergarten management and instruction. For this reason, an additional booklet was developed to deal with kindergarten instruction and classroom management. Coaching was also added to the training program to enhance the transfer of training.

In the second study the staff-development program was conducted by teacher trainers and school counsellors in five locations in the Netherlands. Two months prior to the actual start of the training, the teacher trainers and school counsellors were trained by members of the Department of Educational Sciences to secure the same program implementation as in the first study. In the school year 1989/90, the staff-development program was followed by 89 teachers of grades one through six from 12 schools. From this group, 28 teachers were then selected for participation in the observational study (treatment group). In addition to the training, 18 teachers from this treatment group also received in-class coaching. Fourteen teachers from 6 schools in the same areas were selected for study but did not receive training (control group).

In both of the short-term improvement studies, the contents of the training were the same. The kindergarten teachers in the second study received an additional booklet.

Directly before and after training, the teachers who participated in the observational part of the short-term improvement studies were observed during two mathematics and two reading/language lessons. Following five to seven three-hour workshops, the teachers in the multi-grade classes implemented self-designed plans to increase specific teaching behaviours and pupil time-on-task. Feedback was provided before the start of the first workshop -- based on the results of pretest observation, and after the last training session -- based on the results of posttest observation. This feedback contained information about the time-on-task rates in the classes, the observed instruction- and classroom-management skills, and other aspects of the lessons. Between the workshops, the teachers were also asked to experiment with some of the teaching recommendations in their classrooms.

The design of the training process was guided by the recommendations of Joyce and Showers (1980, 1988). The five major components of training were: 1) presentation of theory; 2) modelling or demonstration; 3) practice; 4) structured feedback; and 5) coaching. The theory was presented in the handbook. Modelling or demonstration of the suggested teaching skills was done using video-fragments and the presentation of case studies in the handbook. Practice under simulated conditions was achieved by role-playing with peers; practice under real conditions was achieved by asking the teachers to experiment with new ideas or improvement plans and report what they had done at the next workshop. As already mentioned, feedback was provided both before and after training.

Based on the pre- and post-training observations, the first improvement study revealed a significant treatment effect for the time-on-task levels of the pupils in the multi-grade classrooms, effective instruction, and the classroom organization and management behaviours of the teachers (Veenman, Lem, & Roelofs, 1989). The second improvement study also revealed a significant treatment effect for the time-on-task levels of the pupils in the multi-grade classes along with the instructional and classroom-management skills of the teachers. Two coaching effects were found, namely for the effective organization of instruction and for dealing with disturbances. The time-on-task levels improved more strongly in classes with coached teachers. The effects on the instructional and classroom-management skills of the teachers and the on-task behaviour of the pupils in the second study were found to be smaller than those in the first study (Roelofs, Veenman, & Raemaekers, 1994).

Research Questions

In the present study, the long-term effects of a staff-development program in school settings with multi-grade classes are examined. The long-term effects of coaching in addition to the staff-development program are also evaluated. The research questions that guided the study were the following: Do teachers who followed the staff-development program in the school years 1986/87 and 1989/90 still use the target behaviours after two and five years of training? Does the training appear to have a lasting effect on the time-on-task levels of the pupils? Are the effects of training greater for teachers who received coaching in addition to participation in the staff-development program? Does the training appear to have a positive effect on pupil achievement?

Methods

Design

The study was designed as a quasi-experiment with two treatment groups (uncoached teachers (n=10) and coached teachers (n=8)) and one control group (n=11). The classrooms in the retention study were selected from the first and second school-

improvement studies.

Subjects

The selection of the teachers for the first and second improvement studies is described in Veenman, Lem, and Roelofs (1989) and in Roelofs, Veenman, and Raemaekers (1994). In the following, the selection of the teachers for the retention study is described.

Of the 17 trained teachers who participated in the observational part of the first improvement study in the school year 1986/87, 8 were willing to participate in the retention study five years later. Of the 10 trained but uncoached teachers who participated in the observational part of the second improvement study in the school year 1989/90, 2 were willing to participate in the retention study two years later. This produced a total of 10 trained but uncoached teachers for follow-up study. Of the 18 coached teachers in the second improvement study, 8 were willing to participate in the retention study two years later. The total of 18 trained teachers (coached or uncoached) came from grades one through six in 9 schools.

For the control group, 11 teachers were recruited from 7 schools with socioeconomic backgrounds and geographic locations comparable to those of the schools in the treatment group (6 teachers from the 9 control-group teachers in the first improvement study and 5 teachers from the 14 control teachers in the second improvement).

In order to test for any self-selection effects among the teachers from the initial improvement studies, the instructional- and classroom-management pretest scores for the teachers who had volunteered for participation in the retention study were compared with the pretest scores for the teachers who had not volunteered for participation. For the treatment groups (n=18), no significant differences were found between the previously participating teachers and the newly recruited teachers. For the control group (n=11), one significant difference was found. The newly-recruited control teachers were found to score significantly lower than the previously-participating teachers on one of the five classroom-management subscales, namely 'adjusting instruction to the needs of the pupils' (p <.05). On the basis of these results, it was concluded that the degree of self-selection was minimal.

Although all of the teachers who participated in the observational part of the first and second improvement studies were asked to participate in the retention study, the number who volunteered was not very impressive (40% for the treatment group and 57% for the control group. It should be noted that the schools and teachers were willing to participate in the first and second improvement studies as they received extensive training in return for their cooperation. Cooperation in a follow-up study two-to-five years later, however, yielded no immediate profit for the schools and the teachers. This was the main reason for not participating in the retention study. Other reasons were: too busy, illness, and the teachers having left the school.

Instrumentation

The instruments used to measure the quantity and the quality of program implementation and the time-on-task levels of the pupils were identical to those used in the first and second improvement studies. These included an observation instrument and a classroom rating-scale. Standardized achievement tests were also administered to measure pupil progress.

Time-on-task and instructional-skills observation.

Observational data on the time-on-task levels for the pupils were collected using a 'predominant activity' time-sampling procedure (Tyler, 1979). To obtain information on the behaviours of both the teachers and the pupils, a predetermined observational sequence was established. The observer examined the behaviour of the first pupil and that of the teacher for seven seconds and then recorded this information in the next thirteen seconds. After observation of all of the pupils, the observer started again with the first pupil.

An observation period lasted 40 minutes with optical and auditory signals (produced by an observation-timer) to indicate the start of the observation period and the start of the coding intervals. The following four pieces of information were recorded for each observation period: (a) the response of the pupil to the task (e.g., on-task, off-task); b) the target group of the teacher (e.g., grade level 5 or 6); c) the task-related activities of the teacher (e.g., supervision, guided practice); and d) the setting of the learning activities for

each grade (e.g., group instruction, seatwork). The observation instrument, entitled COMMIT, included 20 categories. The most important observational variables used in the retention study are listed in Tables 1 and 2. The observational data were collected using paper-and-pencil forms that could then be optically scanned.

Prior to the collection of the observational data, the three observers went through a training program of about 40 hours. This involved the coding of videotapes as well as live coding. The training was conducted by an experienced trainer from the second improvement study who had been trained by an observer from the first improvement study to secure identical observational procedures and coding. The inter-observer reliability checks, estimated using analysis of variance (Winer, 1971), ranged from .70 to 1.00 (median .97) with the exception of one category: 'target group whole class.' The inter-observer reliability for this category was found to be .25, which may be due to the fact that the teachers observed in the multi-grade setting rarely directed their teaching to the entire class. All of the classrooms were observed by at least two different observers in order to minimize any observer effects.

Classroom rating scale.

After each observation, the Management and Instruction Scale (MIS) was completed by the observer to assess teacher and pupil behaviours. The assessment consisted of five-point scales concerned with instructional skills, lesson design and execution, management of pupil behaviour, classroom organization, and the level of disruptive or inappropriate behaviour.

The items in the MIS are based on the research of Evertson et al. (1983), Good, Grouws, and Ebmeier (1983), and Rosenshine and Stevens (1986). The MIS contains 31 items and five subscales: (1) instructional skills, (2) organizing instruction, (3) use of materials and space, (4) adjusting instruction, and (5) dealing with disturbances. The alpha-coefficients of the internal reliability of the different subscales ranged from .66 to .91 (see Table 2). The inter-observer reliability checks for all of the subscale-scores, estimated using analysis of variance, ranged from .63 to .81 (median .74).

Achievement tests.

Three standardized achievement tests were used to measure pupil progress: one for decoding or technical reading (Brus-Voeten Test), one for reading comprehension (CITO and Aarnoutse Test), and one for mathematical skills (De Vos Test). These tests were administered in the classes with trained and the classes with untrained teachers (grades one through six) to determine the effects of the staff-development program on pupil achievement. The technical reading and mathematical tests were suitable for all of the grades involved. For reading comprehension, different tests that nevertheless measured the same construct were used for each grade.

Data Collection

For the first improvement study, the pretest took place in November-December 1986 and the posttest in May-June 1987. For the second improvement study, the pretest was administered in November-December 1989 and the posttest in May-June 1990. The follow-up test was administered during the period May-July 1992, five years after the posttest in the first improvement study and two years after the posttest in the second improvement study.

Each teacher in the retention study was observed for one mathematics and one reading/language lesson. All of the observations took place in the morning. The observational data of the COMMIT were then expressed in minutes. The teacher and pupil behaviours within each category were averaged for each class and each teacher, and the observations for a particular subject (mathematics and reading/language) were then averaged to produce the mean rates for each observation period (i.e., the retention data). It was recognized that the observational variables were not independent of each other and that the coding of an event into one category excludes it from inclusion in all of the other categories for that interval.

Subscale scores for the MIS were computed by adding the values of the responses for each subscale together.

In testing for differences between the treatment teachers and the control teachers, a significance level of 5% was used (one-tailed). The class or teacher was the unit of

analysis for the observational data. The pupil was the unit of analysis for the achievement data. For a more detailed description of the design, instrumentation, and data collection in this study, see Raemaekers and Veenman (1994).

Results

To determine the effects of the staff-development program, an analysis of covariance (ANCOVA) with the pretest as a covariate was undertaken. The posttest and retention scores constituted the dependent variables with treatment (training or no training) as an independent variable. Three comparisons were made: (1) the control group versus the treatment group; (2) coaching versus no coaching; and 3) the treatment group from 1986/87 versus the treatment group from 1989/90. It should be noted that only the teachers who did not receive coaching in the second improvement study could be used for comparison to the teachers in the first improvement study (which did not include coaching).

When comparing the treatment group (i.e., trained teachers either with or without coaching) with the control group to test for initial differences, no significant differences were found. The treatment groups with and without coaching differed at pretest for the observational category of 'procedural activities'. The pupils in the classes with trained teachers who also received coaching spent less time on procedural activities than pupils in the classes with trained teachers who did not receive coaching ($p < .05$). In addition, the coached teachers were rated significantly higher than the non-coached teachers on one of the subscales of the MIS: 'use of materials and space' ($p < .05$). In sum, no significant differences were found between the control group and the treatment groups prior to training. Minor differences were found between the two treatment groups (i.e., the coached versus uncoached teachers).

Training Effects

A summary of the descriptive statistics for each independent variable on the COMMIT and MIS is presented in Table 1. In Table 2, the results of the statistical tests are summarized. In the columns regarding the 'training effects' of the staff-development

program, representing the differences between the treatment and the control groups and the treatment groups themselves, the post- and retention-test scores have been averaged to produce a combined score. This score is then compared for the experimental versus control groups and the different treatment groups with the pretest score as a covariate. The results of the significance tests are expressed as *t*-values (expressing the differences between two contrasting groups), and the information found in Table 1 should be kept in mind when interpreting these results.

Effects of the training program.

A comparison of the mean scores for the treatment group (coached and uncoached teachers) with the mean scores for the control group showed the staff-development program to have a significant effect the time-on-task rates for the pupils. The treatment-group pupils exhibited higher time-on-task levels than the control-group pupils: 80% (32.0 minutes) versus 72% (28.7 minutes), which proved to be statistically significant ($p<.01$). The treatment-group pupils spent significantly less time waiting for the teacher and were more engaged in their work than the control-group pupils.

Table 2 also presents the outcomes regarding the amount of time-on-task during class instruction and individual seatwork. The treatment-group pupils were in both settings significantly more on-task than the control-group pupils. During class instruction, the treatment-group pupils were found to spend 86% of the observation time on their learning tasks (34.4 minutes), and during individual seatwork 78% of the observation time (31.0 minutes). The respective figures for the control-group pupils were 77% (30.9 minutes) and 66% (26.6 minutes). These results indicate that the trained teachers were able to establish classes with a greater proportion of the pupils engaged in learning tasks (on-task) than in the control classes.

In Table 2, information regarding some of the teacher behaviours is summarized to estimate the degree of program implementation. Significant differences between the treatment and control teachers were found for the variables 'review of previous work' and 'no teaching behaviour.' The treatment teachers spent significantly more time than the control teachers on behaviours intended to activate pupils' prior knowledge of the subject matter and significantly less time than the control teachers on organization of the

classroom (5% versus 3% and 1% versus 6%).

Finally, the results for the Management and Instruction Scale (MIS) are also summarized in Table 2. The results show those teachers who participated in the training to attain substantially higher scores than the control teachers. Significant implementation effects were found for four of the five subscales. No treatment effect was found for the subscale 'use of materials and space.' In general, the trained teachers were found to use more effective instructional, classroom- management, and organizational techniques than the control teachers. The trained teachers improved markedly on instructional skills, the organization of instruction, the adjustment of the instruction to pupil abilities, and dealing with disturbances.

Effects of coaching.

No significant differences were found between the coached and uncoached teachers. Coaching had no lasting effect on time-on-tasks levels of the pupils or the instructional and classroom- management skills of the trained teachers.

Effects of application time.

It was expected that the teachers who had participated in the training five years prior would show higher implemention rates than the teachers who had participated in the training just two years prior. The findings in the column 'training effects' in Table 2 reflect the influence of treatment in 1986/87 versus 1989/90. No significant differences were found between these two groups. It should be noted, however, that the 1989/90 treatment group represents the scores of only two teachers. The results of this test should therefore be interpreted with caution.

Retention Effects

In order to examine the retention effects of the staff-development program, the differences between the retention scores and the posttest scores (expressed as gain scores) were compared for the experimental (coached and uncoached teachers) versus control groups

and the different treatment groups with the pretest score as a covariate. The outcomes of these comparisons are grouped under the heading 'retention effects' in Table 2.

Effects of the training program on retention.

When the pretest scores were controlled for, no significant difference was found between the treatment and control groups. One may conclude that the treatment teachers did not show an increase between the posttest and the retention test in the application of the desired instructional and classroom- management skills. No indication of further growth in the 'executive control' of these skills was found. In other words, it was difficult to improve upon the results achieved directly after training.

Effects of coaching on retention.

No significant differences were found between the uncoached and coached teachers. In spite of the additional support and classroom assistance, the coached teachers did not gain more between the post and retention tests than the uncoached teachers.

Effects of application time on retention.

Three significant differences were found between the 1989/90 treatment teachers and 1986/87 treatment teachers. Although no significant differences were found in the time-on-task levels for the pupils in these two groups, the 1989/90 group showed a significant decrease in the off-task behaviour of 'waiting' ($p < .05$) when compared to the 1986/87 group. The 1989/90 treatment teachers also showed larger gains for 'guided practice' ($p < .05$) and the 'instructional skills' subscale of the MIS ($p < .01$) when compared to the 1986/87 group. Unexpectedly, the 1989/90 treatment teachers gained more on these variables between the post and retention test than the 1986/87 treatment teachers. Considering the number of dependent variables, however, the differences between the two treatment groups appear to be minimal.

Pupil Achievement

Table 3 summarizes the long-term effects of the staff-development program on pupil achievement. It should be noted that the number of teachers who participated in the pupil achievement part of the retention study is larger than the number of teachers who participated in the observational part of the retention study. All of the teachers did not agree to open their classes for follow-up observations. Of the 12 classes in the follow-up treatment group with uncoached teachers, 7 of the teachers had participated in the first improvement study and 5 in the second improvement study. All of the classes in the follow-up treatment group with coached teachers (n = 14) had participated in the second improvement study (coaching was not applied in the first improvement study). In the follow-up control group participated 14 teachers (7 from the first and 7 from the second improvement study).

To account for the differences in the lengths of the reading comprehension and mathematics tests, all of the raw pupil scores were standardized per grade level using z-scores. Analysis of variance was then applied to test for any differences between the treatment and control groups, the coached and uncoached teachers, or the 1986/87 and 1989/90 treatment groups.

No significant cross-grade achievement differences were found. The treatment classes did not perform better than the control classes. Systematic differences were also not found per grade level (see Raemaekers & Veenman, 1994). Furthermore, no significant differences were found between the classes of the coached and uncoached teachers or between the classes of the teachers in the 1986/87 treatment group and the classes of the teachers in the 1989/90 treatment group. Note that the last comparison involved only uncoached teachers. These findings should also be treated with caution as pre-treatment achievement data were not available. This means that any initial differences in pupil achievement could not be controlled for.

Discussion

In two previous studies, the short-term effects of the staff-development program 'Dealing with Multi-Grade Classes' were assessed. The results of these studies showed important

gains in instructional skills and the way in which the trained teachers organized instruction and adapted it to the pupils. Classroom-management skills such as the use of materials/space, and dealing with disturbances also improved markedly for the trained teachers. The time-on-task levels for the pupils with trained teachers also increased substantially. Finally, coached teachers were found to differ from uncoached teachers on only two aspects of instruction and classroom management: organization of the instruction and dealing with disturbances. For both aspects, the coached teachers showed larger gains than the uncoached teachers.

In the present study, the long-term effects of the staff-development program were examined. Observational data collected two and five years after the conclusion of the training program showed the trained teachers to still demonstrate the target teaching behaviours. The group differences two and five years after training show the staff-development program to have enhanced the skills of teachers in multi-grade classes. The target skills appear to have been transferred and sustained over time.

A number of the effects of the training are outlined in Figure 1 as transfer patterns or transfer curves. These patterns illustrate the changes between pre-, post-, and retention-test. The tranfer patterns with regard to the time-on-tasks levels and the review of previous work demonstrate a curve described by Den Ouden (1992) as reflective of a "long-term change process". The transfer pattern with respect to guided practice indicates a failure to maintain the post-training level. No implementation and no transfer are seen for the skill of monitoring, which suggest that the teachers have no intention of changing their behaviour in the desired direction. The transfer patterns for the instructional skills and organization of the instruction reflect the outcomes of the Management and Instruction Scale (MIS). All five subscales of the MIS demonstrate a "long-term change process curve." For one subscale, namely the use of materials and space, the control teachers nevertheless changed their behaviour more in the desired direction than the treatment teachers between post- and retention test (see Table 1).

A number of transfer strategies were applied both before and during the training period. The teachers were briefed on the importance of the training objectives, content, process, and application for the teaching of multi-grade classes. Both trainers and trainees were involved in the planning of the program. The trainees were selected because of their desire to improve their instructional and classroom-management skills for the teaching of

multi-grade classes. Baseline performance data were used during the training program to formulate clear improvement plans. The training was systematically developed and based on the components of effective training outlined by Joyce and Showers (1988). Joint expectations for improvement were established. Opportunities to practice newly acquired skills were provided in both simulated and real classroom-settings. Realistic work-related tasks and cases were provided (situated cognition). A coaching component was added to the program in order to facilitate the transfer of newly acquired skills to the classroom. Opportunities for the discussion of ideas and their applications were created. Case studies were used to illustrate how other teachers had implemented a particular teaching skill in a multi-grade classroom. Trainees planned and then discussed their plans. The incorporation of strategies to facilitate the transfer of training (cf. Broad & Newstrom, 1992) may have contributed to the lasting effects of the staff-development program.

One strategy for the promotion of transfer did not appear to be particularly successful, however. The coaching of teachers did not produce a lasting effect on the on-task levels of the pupils or the skills of the teachers. This finding may be explained by the way in which coaching was performed. The results of the second improvement study (see Roelofs, Veenman, & Raemaekers, 1994) suggest that all of the coaching functions may not have been implemented effectively. The provision of support and technical feedback with regard to the newly acquired skills was found to be valued positively by the coached teachers. Little attention was paid to how the newly acquired skills should actually be applied in the classroom or adapted to the characteristics of the pupils. Joyce and Showers (1988) nevertheless argue that these last two functions are particularly important for the integration of the newly acquired behaviours into the daily teaching repertoire. In this respect, critical differences between the coaches were observed and a more structured set of guidelines for the coaching of teachers on the job may be needed.

No effects were found for the length of application. The teachers who had participated in the training five years prior showed no higher implemention rates than the teachers who had participated in the training two years prior. The 1989/90 treatment group contained only two teachers, which suggests that the possibility of greater skill perfection over time should be evaluated in further research.

The full transfer of training in the present context means that the level of skill should increase beyond the level demonstrated at the end of the staff-development

program. Such an increase was not found in the present study. When compared to control teachers, the treatment teachers did not show an increase over time in the application of the desired instructional and classroom-management skills. Interviews with the teachers using the Levels of Use of the Innovation (LoU) (Hall & Loucks, 1977; Hord, Rutherford, Huling-Austin, & Hall, 1987) showed the trained teachers to be functioning at level 3, 'mechanical use', and level 4a, 'routine use' (see Raemaekers & Veenman, 1994). A 'stable, routine' pattern of use appeared to have been established. This may be due to the fact that most of the instructional and classroom-management skills were defined in very concrete teaching terms (e.g., review previous work, provide individualized help). Such concrete behavioural descriptions provide little room for further refinement or higher frequencies of application. Each lesson, for example, has only one moment that is appropriate for the review of previous work; teachers usually review the previous work at the beginning of a lesson. The provision of individualized help depends on the number of pupils in need of help. No teacher will deliberately create confusing situations in order to maximize the number of opportunities for individualized help.

It is interesting to note that although the trained teachers achieved higher time-on-task levels than the untrained teachers, these higher time-on-task levels did not result in higher pupil achievement. No significant achievement differences were found for the classes with trained versus untrained teachers. There are two potential explanations for why more time-on-task was not associated with higher pupil achievement. First, the teachers may have treated time as a homogeneous entity. More time may simply have been taken to mean more of the same. No effort was made in this study to partition time into various pupil or teacher behaviours. We do not know therefore, if more time was spent on the right tasks. Time was measured quantitatively and not qualitatively. In a revised edition of the staff-development program, the teachers should be trained to examine their time-on-task data in light of the question: time on *what* task? The quality of the task may determine just which and how much learning occurs. Second, the staff-development program was mainly directed at the improvement of teacher behaviours in multi-grade classes. A stronger coupling between teacher and pupil behaviours may be needed. To improve pupil learning, teachers may need to be stimulated to identify the desired pupil behaviours and then the teacher behaviours needed to evoke such pupil behaviours. In such a way, the time-on-task levels of the pupils may become more directly

related to their achievement.

Two to five years after training, the guided practice aspect of the staff-development program appears to have been wiped out. Guided practice is considered an important teaching skill in multi-grade classes because it shows whether the pupils are prepared to start individual seatwork or not. If the pupils are not ready, errors and frustration may decrease their interest and their performance. In addition, the pupils in the first group may interrupt the teacher during the instruction of the second or third groups, which may produce confusion and thereby decrease pupil performance further. Only when pupils understand the new material and are largely able to work out the problems correctly can the teacher proceed to the instruction of other groups in a multi-grade class. As the pretest data suggest (see Table 1), teachers often simply assume that their presentations have been understood. After their presentations, they switch directly to seatwork or the independent work portion of the lesson, and leave the instructed group without supervision to instruct a new group. Directly after training, the teachers in this study were found to use guided practice significantly but not impressively more than before training. Two or five years after training, however, this treatment effect disappeared. The teachers attempted to use the skill for a period of time, but its use quickly diminished to the level of the pre-training baseline. This decline may be due to a perceived lack of success with the skill, the constraints of the complex work environment associated with a multi-grade class, a lack of support for the use of this particular skill, or a combination of these factors. In the future, greater attention should be paid to transfer strategies in the work-environment of teachers in multi-grade classes (e.g., follow-up support).

Acknowledgements

The research reported here was supported by a grant from the Institute for Educational Research in the Netherlands (SVO-project 1111). The authors gratefully acknowledge the following persons for their contributions to this study: Rinus Voeten, Cora van den Berg, Mirjam van Moorsel, Carla Motshagen, Desiree Jacobs, Nelleke Hofs, and the participating principals and teachers for their cooperation.

REFERENCES

Baldwin, T.T., & Ford, J.K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41(1), 63-105.

Broad, M.L., & Newstrom, J.W. (1992). Transfer of training: Action-packed strategies to ensure high payoff from training investments. Reading: Addison-Wesley.

Bennett, B. (1987). The effectiveness of staff development training practices: A meta-analysis. Unpublished doctoral dissertation, University of Oregon.

Den Ouden, M.D. (1992). Transfer na bedrijfsopleidingen [Transfer after corporate training]. Amsterdam: Thesis Publishers.

Evertson, C.M., Emmer, E.T., Sanford, J.P., & Clements, B.S. (1983). Improving classroom management: An experiment in elementary school classrooms. Elementary School Journal, 84(2), 173-188.

Evertson, C.M., Emmer, E.T., Clements, B.S., Sanford, J.P., & Worsham, M.E. (1984). Classroom management for elementary teachers. Englewood Cliffs, N.J.: Prentice Hall.

Fraser, B.J., Walberg, H.J., Welch, W.W., & Hattie, J.A. (1987). Syntheses of educational productivity research. International Journal of Educational Research, 11(147-252).

Gayfer, M. (1991). The multi-grade classroom: Myth and reality. A Canadian study. Toronto: Canadian Education Association.

Commissie Evaluatie Basisonderwijs. (1994). Inhoud en opbrengsten van het basisonderwijs [Contents and results of primary education]. Leiden: Distributiecentrum DOP.

Good, T.L., & Brophy, J.E. (1986). School effects. In M.C. Wittrock (Ed.), Handbook of Research on Teaching (pp. 570-602). New York: Macmillan.

Good, T.L., Grouws, D.A., & Ebmeier, H. (1983). Active mathematics teaching. New York: Longman.

Hall, G.E., & Loucks, S.F. (1977). A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, 14(3), 263-276.

Her Majesty's Inspectorate. (1982). The new teacher in school. London: HMSO, Department of Education and Science.

Hord, S.M., Rutherford, W.L., Huling-Austin, L., & Hall, G.E. (1987). Taking charge of change. Alexandria: Association for Supervision and Curriculum Development.

Joyce, B.R., & Showers, B. (1980). Improving inservice training: The messages of research. Educational Leadership, 37(6), 379-385.

Joyce, B.R., & Showers, B. (1988). Student achievement through staff development. New York: Longman.

Kierstead, J. (1986). How teachers manage individual and small-group work in active classrooms. Educational Leadership, 44(2), 22-25.

Kounin, J.S. (1970). Discipline and group management in classrooms. New York: Holt, Rinehart & Winston.

Levine, D.U., & Lezotte, L.W. (1990). Unusually effective schools. Madison: National Center for Effective Schools Research & Development.

Raemaekers, J., & Veenman, S. (1994). Institutionalisering van teamgerichte nascholing: Duurzame effecten van de cursus Omgaan met Combinatieklassen [Institutionalisation of a staff-development program: Long-term effects of the program Dealing with Multi-grade Classes]. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

Roelofs, E., Veenman, S., & Raemaekers, J. (1994). Improving instruction and classroom behaviour in mixed-age classrooms: Results of two improvement studies. Educational Studies, 20(1), 105-126.

Rosenshine, B.V., & Stevens, R. (1986) Teaching functions. In M.C. Wittrock (Ed.), Handbook of research on teaching (376-391). New York: MacMillan.

Tyler, S. (1979). Time-sampling: A matter of convention. Animal Behavior, 27, 801-810.

Van Tulder, M. (1992). Nascholing en onderwijsvernieuwing [In-service education and educational innovation]. Doctoral dissertation, University of Nijmegen. Nijmegen: Universiteitsdrukkerij.

Veenman, S., Lem, P., & Roelofs, E. (1989). Training teachers in mixed-age classrooms: Effects of a staff-development programme. Educational Studies, 15(2), 165-180.

Veenman, S., Lem, P., & Nijssen, F. (1988) Omgaan met combinatieklassen: Een programma voor schoolverbetering [Dealing with multi-grade classes: A staff-development program for school improvement]. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs.

Veenman, S., Lem, P., Voeten, M., Winkelmolen, B., & Lassche, H. (1986) Onderwijs in combinatieklassen [Instruction in multi-grade classes]. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs.

Veenman, S., Voeten, M., & Lem, P. (1987) Classroom time and achievement in mixed-age classes. Educational Studies, 13(1), 75-89.

Walberg, H.J. (1986). Synthesis of research in teaching. In M.C. Wittrock (Ed.), Handbook of Research on Teaching (pp. 917-924). New York: Macmillan.

Walsh, K., Dunne, R., Stoten, B., & Stewart, J.D. (1984). Falling school rolls and the management of the teaching profession. Windsor: NFER-Nelson.

Winer, B.J. (1971). Statistical principles in experimental design. New York: McGraw-Hill.

Table 1. Means of observation categories of the COMMIT and mean rates of the Management and Instruction Scale (MIS).

	Treatment Group (n=18)															Control Group (n=11)		
	Coached (n=8)			Uncoached (n=10)									Total (n=18)					
	1989/90 (n=8)			1986/87 (n=8)			1989/90 (n=2)			Total (n=10)								
Variables	Pre	Post	Ret	Pre	Post	Ret	Pre	Post	Ret	Pre	Post	Ret	Pre	Post	Ret	Pre	Post	Ret
COMMIT																		
Pupil behaviour																		
On-task	28.1	30.8	32.6	28.0	33.0	31.3	29.4	31.2	33.9	28.3	32.7	31.8	28.2	31.9	32.1	28.7	27.3	30.1
Procedural	2.8	2.3	2.5	5.0	3.3	2.8	2.7	2.2	2.3	4.5	3.1	2.7	3.8	2.7	2.6	3.4	3.7	2.7
Waiting	2.1	1.3	0.8	1.0	0.4	0.9	1.7	1.5	0.3	1.1	0.6	0.8	1.6	0.9	0.8	1.4	1.4	1.1
Not engaged	7.0	5.6	4.2	6.0	3.3	5.1	6.3	5.1	3.5	6.1	3.7	4.8	6.5	4.5	4.5	6.6	7.5	6.1
Setting and on-task																		
Instructon and on-task (%)	71.8	81.6	86.0	76.7	86.6	89.4	77.2	82.3	90.1	76.8	85.7	89.6	74.6	83.9	88.0	75.4	72.7	81.8
Seatwork and on-task (%)	69.6	75.4	77.9	68.3	81.4	75.6	70.8	76.0	78.8	68.8	80.3	76.2	69.2	78.1	77.0	69.6	63.5	69.5
Teacher behaviour																		
Review of previous work	0.1	2.2	3.3	0.5	1.9	0.6	0.0	2.8	2.5	0.4	2.1	1.0	0.3	2.1	2.0	0.2	0.4	0.4
Guided practice	0.9	1.0	0.0	2.9	6.3	0.4	0.0	0.8	1.6	2.3	5.2	0.6	1.7	3.3	0.4	2.0	0.9	0.9
Monitoring	5.8	4.6	4.3	4.0	3.9	3.4	5.1	5.7	0.9	4.2	4.3	2.9	4.9	4.4	3.5	3.2	4.0	3.0
Transitions	5.8	5.6	4.0	6.0	5.4	3.6	4.3	3.8	1.3	5.6	5.1	3.1	5.7	5.3	3.5	4.6	5.5	3.6
No teaching behaviour	1.3	1.0	1.2	3.0	1.1	2.2	1.0	1.5	0.2	2.6	1.2	1.8	2.0	1.1	1.5	2.2	3.3	1.8
Subscales MIS																		
Instructional skills	22.4	25.1	27.5	24.3	29.0	25.4	21.0	25.0	30.3	23.7	28.2	26.4	23.1	26.8	26.9	22.8	22.0	23.6
Organising instruction	17.5	20.4	22.4	16.8	20.7	20.9	18.4	20.8	22.5	17.1	20.7	21.2	17.3	20.6	21.7	17.8	18.0	18.7
Use of materials and space	24.6	27.5	28.8	21.1	25.2	25.4	25.5	26.3	29.5	22.0	25.4	27.0	23.2	26.3	27.8	24.0	24.6	27.8
Adjusting instruction	21.4	24.1	26.3	20.7	24.4	24.8	22.0	24.3	26.5	21.0	24.4	25.2	21.2	24.3	25.7	21.5	21.3	23.4
Dealing with disturbances	20.4	24.2	26.3	19.7	24.6	23.1	20.1	24.6	27.5	19.8	24.6	24.0	20.1	24.4	25.0	19.8	20.2	22.0

Note: Mean frequencies of pupil and teacher behaviour are expressed in minutes (lesson period = 40 minutes). Ret = retention test.

Table 2. Training and retention effects: Results of t-tests on adjusted mean scores of observation categories of the COMMIT and subscales of the Management and Instruction Scale (MIS) of the rentention study (pretest scores used as covariates)

	Training effects			Retention effects		
	Training: exp. vs. con.	Coaching: coached vs. uncoached	Time: 1986/87 vs. 1989/90	Training: exp. vs. con.	Coaching: coached vs. uncoached	Time: 1986/87 vs. 1989/90
Observation categories	*t*	*t*	*t*	*t*	*t*	*t*
COMMIT						
Pupil behaviour						
On-task	3.83**	-0.28	-0.19	-1.29	0.69	-1.60
Procedural	-1.69	0.24	-0.10	1.15	0.38	-0.32
Waiting	-2.19*	-0.20	0.12	-0.11	-0.31	2.08*
Not engaged	-3.14**	0.26	-0.01	1.07	-1.19	1.84
Setting and on-task						
Instructon and on-task (%)	3.44**	-0.36	0.41	-0.80	-0.14	-0.52
Seatwork and on-task (%)	3.86**	-0.37	0.39	-1.97	1.02	-1.44
Teacher behaviour						
Review of previous work	4.93**	1.78	-2.05	-0.26	1.65	-0.60
Guided practice	1.56	-1.75	0.68	-1.47	0.91	-2.56*
Monitoring	-0.25	0.54	0.43	-0.45	0.96	1.22
Transitions	-1.52	1.11	0.89	-0.09	0.37	0.27
No teaching behaviour	-2.59*	0.27	-0.09	1.52	0.29	1.28
Subscales MIS						
Instructional skills (α=.90)	4.89**	-0.88	-0.87	-0.27	1.04	-3.75**
Organizing instruction (α=.84)	4.10**	0.22	-0.03	0.56	0.73	-0.71
Use of materials and space(α=.66)	1.95	1.00	-0.24	-0.83	-0.46	-0.62
Adjusting instruction (α=.86)	3.37**	0.24	-0.22	-0.48	0.69	-0.91
Dealing with disturbances (α=.91)	3.12**	0.09	-0.90	-0.51	0.78	-1.62

Note: Exp. = experimental (treatment) group; Con. = control group. Time = application time.
* = p<.05; ** = p<.01

Table 3. Means, standard deviations, and ANOVA-test results for *z*-scores concerning achievement outcomes across grades

Subject area	Treatment group			Control group				
	N	Mean	SD	N	Mean	SD	*F*	*p*
Technical reading	630	0.01	1.00	349	-0.02	0.99	0.21	.64
Reading comprehension	634	0.04	1.01	329	-0.07	0.97	2.63	.11
Mathematics	632	-0.02	1.00	337	0.04	1.00	0.90	.34

Note: Treatment group $N = 26$ classes, control group $N = 14$ classes.

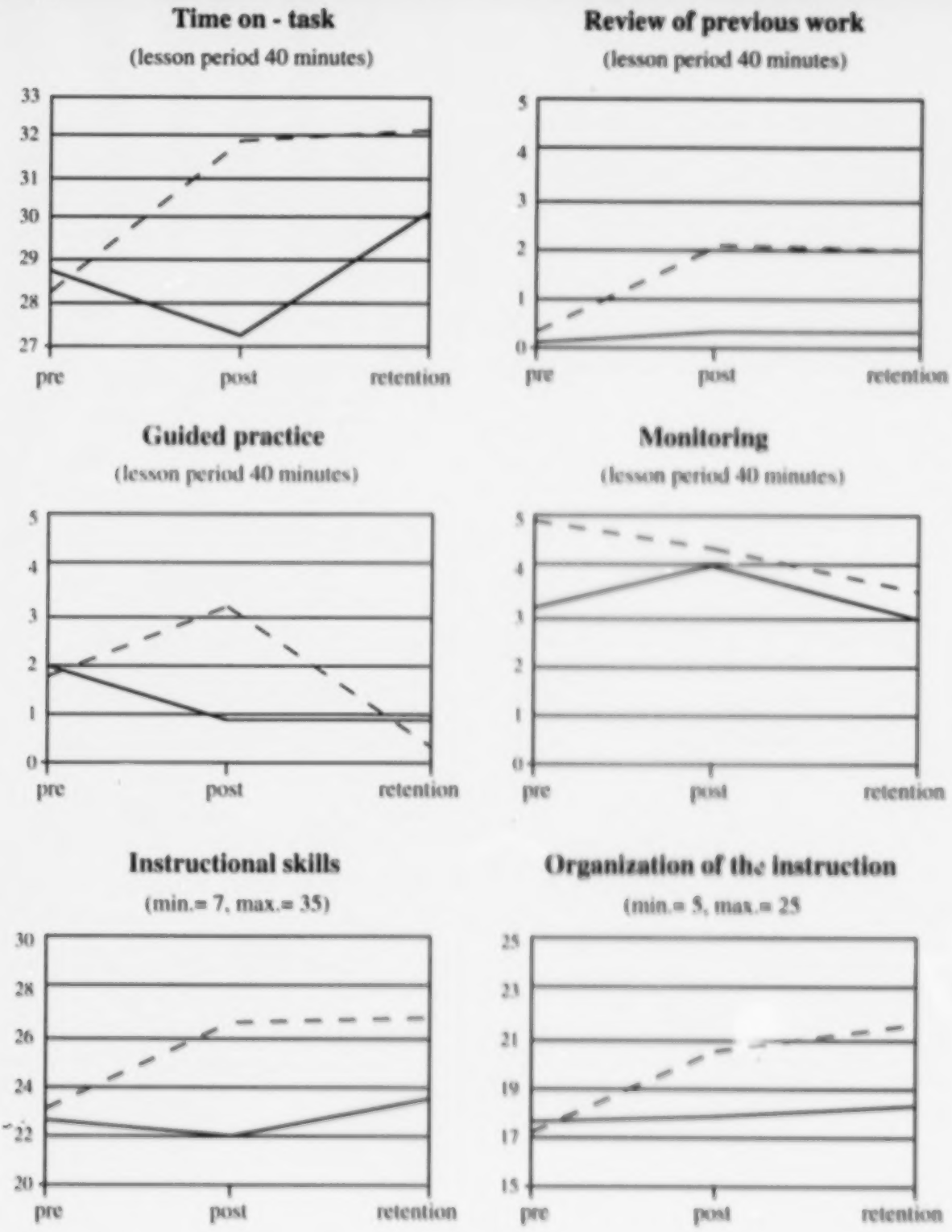

Figure 1. Examples of transfer patterns.